



Western, L. M., Rougier, J., Watson, I. M., & Francis, P. (2020). Evaluating nonlinear maximum likelihood optimal estimation uncertainty in cloud and aerosol remote sensing. *Atmospheric Science Letters*, [e980]. <https://doi.org/10.1002/asl.980>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1002/asl.980](https://doi.org/10.1002/asl.980)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley Open Access at <https://doi.org/10.1002/asl.980> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

RESEARCH ARTICLE

Atmospheric Science Letters

RMETS

# Evaluating nonlinear maximum likelihood optimal estimation uncertainty in cloud and aerosol remote sensing

Luke M. Western<sup>1,2</sup>  | Jonathan C. Rougier<sup>3</sup> | I. Matthew Watson<sup>1</sup> | Peter N. Francis<sup>4</sup>

<sup>1</sup>School of Earth Sciences, University of Bristol, Bristol, UK

<sup>2</sup>Atmospheric Chemistry Research Group, School of Chemistry, University of Bristol, Bristol, UK

<sup>3</sup>School of Mathematics, University of Bristol, Bristol, UK

<sup>4</sup>Met Office, Exeter, UK

## Correspondence

Luke M. Western, Atmospheric Chemistry Research Group, School of Chemistry, University of Bristol, Bristol, UK.  
Email: luke.western@bristol.ac.uk

## Funding information

Natural Environment Research Council, Grant/Award Number: NE/K007912/1

## Abstract

Uncertainty estimates are important when retrieving properties of clouds and aerosols from satellites measurements. These measurements must be interpreted using a form of inverse theory, such as optimal estimation. In atmospheric remote sensing these inverse methods often assume that the forward model is linear in the region of uncertainty. This assumption is not necessarily valid. This paper presents an exact confidence procedure in contrast to the linear approximation using a maximum likelihood estimator. Two simple examples of retrieving the effective radius and optical depth of a volcanic ash cloud and water cloud show a discrepancy between the linear approximation and the exact procedure. The exact procedure is especially useful for inference where the entire parameter space has been forward modelled prior to or during the retrieval, such as using look up tables. When the inference method calculates the likelihood over the whole parameter space, it is less computationally expensive than a linear approximation.

## KEYWORDS

aerosol retrievals, cloud layers, confidence intervals, inverse modelling

## 1 | INTRODUCTION

Statistical methods make it possible to detect and quantify parameters of interest describing clouds and aerosols in the atmosphere from observations (e.g., Mackie *et al.*, 2010; Poulsen *et al.*, 2012; Pavolonis *et al.*, 2013; Western *et al.*, 2018). An example of quantifying a parameter could be inferring the abundance of an aerosol from measurements of infrared brightness temperature or reflectance of solar radiance using inverse methods. The method relies on a forward model that uses input parameters to simulate the observation (see for example, Rodgers, 2000). The inverse method finds the inputs for

which the simulated observations most closely match the actual observations according to some metric and ideally some uncertainty in this estimate. A simple metric of closeness is the root-mean-squared error, but could similarly be found by using a statistical model and finding the maximum likelihood estimate. One approach forward-models the entire parameter space before any retrieval takes place, which is sometimes called a look-up table; see, for example, Nakajima and King (1990) on stratiform cloud layers, which is used routinely (e.g., Benas *et al.*, 2019; Peers *et al.*, 2019). Uncertainties in the maximum likelihood estimate are often ignored; where they are not, they are usually approximated using a

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Atmospheric Science Letters* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

linearization of the forward model, which gives rise to elliptical uncertainty regions. This paper shows that for maximum likelihood estimation there exists an exact confidence procedure to quantify uncertainty that does not necessarily produce elliptical regions. We compare the exact confidence region to the elliptical approximation in two simple examples: the retrieval of the effective radius and optical depth of a volcanic ash cloud and water cloud. For the given examples, if the exact confidence region is not elliptical this shows that a linear assumption is insufficient in describing the region of uncertainty.

## 2 | UNCERTAINTY REGIONS

Nonlinear optimal estimation in aerosol and cloud remote sensing, in many cases, relies on a cost function to retrieve  $n$  parameters  $\mathbf{x}$  from  $m$  observations  $\mathbf{y}$ . This needs a forward model  $\mathbf{F}(\cdot)$  to simulate the observations  $\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon$ , where  $\epsilon$  is the model-measurement error.

We treat this as a statistical model by providing a statistical description of the distribution of  $\epsilon$  given  $\mathbf{x}$ . The standard statistical model is that  $\epsilon$  is Normal with expectation zero and (specified) variance  $\mathbf{S}_\epsilon$ , independently of  $\mathbf{x}$ . Under this statistical model the likelihood function  $L(\cdot)$  for  $\mathbf{x}$ , where we maximize the probability associated with observations  $\mathbf{y}$ , has the form

$$-2\ln L(\mathbf{x}) := -2\ln p(\mathbf{y} | \mathbf{x}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + c, \quad (1)$$

where  $c$  is a constant that does not depend on  $\mathbf{x}$ . The likelihood function evaluates the probability of  $\mathbf{x}$  when the true observation is  $\mathbf{y}$ . In a “less statistical” approach, the right-hand side is interpreted as a cost function to be minimized in  $\mathbf{x}$ . Minimizing the cost function is equivalent to finding the maximum likelihood estimate,  $\hat{\mathbf{x}}$ . The advantage of a statistical approach is that it provides an assessment of uncertainty about  $\mathbf{x}$ , as we now describe.

A set estimator for  $\mathbf{x}$  is a function  $\mathbf{y} \mapsto C(\mathbf{y})$  where  $\mathbf{y}$  are the observations, and  $C(\mathbf{y})$  is a set of possible values for  $\mathbf{x}$ .  $C$  is a “level  $(1 - \alpha)$  confidence procedure” exactly when

$$\Pr\{\mathbf{x} \in C(\mathbf{y}); \mathbf{x}\} \geq (1 - \alpha)$$

for all  $\mathbf{x}$ , where  $\mathbf{y}$  is treated as a vector of random quantities, parameterized by  $\mathbf{x}$ . The probability on the left-hand side is termed the “coverage” of  $C$  at  $\mathbf{x}$ .  $C$  is “exact” if its coverage is  $(1 - \alpha)$  for all  $\mathbf{x}$ ; otherwise it is “conservative”, meaning the coverage may be larger than  $(1 - \alpha)$

for some  $\mathbf{x}$ . Confidence sets (or confidence intervals for scalar quantities) are a standard approach to quantifying uncertainty. For more on confidence procedures, see Casella and Berger (2002, ch. 9); for a critical appraisal, see Morey *et al.* (2016).

Exact confidence procedures are rare, and do not exist for most statistical models. However, in our case it is straightforward to derive an exact confidence procedure for  $\mathbf{x}$ . The statistical model implies that

$$[\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] \sim \chi_m^2, \quad (2)$$

where  $\chi_m^2$  is the chi-squared distribution with  $m$  degrees of freedom, the number of observations (see for example, Mardia *et al.*, 1980, Ch. 2). This converts directly into an exact level  $(1 - \alpha)$  confidence procedure for  $\mathbf{x}$ ,

$$C(\mathbf{y}) = \left\{ \mathbf{x} : [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] \leq \chi_m^{-2}(1 - \alpha) \right\}, \quad (3)$$

where  $\chi_m^{-2}$  is the quantile function of the chi-squared distribution with  $m$  degrees of freedom. The confidence procedure  $C$  is a level set of the likelihood function, and thus has various attractive properties, including being transformation-invariant. Transformation-invariance means that the transformed set retains its properties as a confidence procedure. For example, if  $\mathbf{x}_i$  was cloud top height in metres, and we computed a 95% confidence set for  $\mathbf{x}_i$ , and then transformed each element of the set into atmospheric pressure, then the resulting set would be a 95% confidence set for atmospheric pressure.

We are often interested in individual components of  $\mathbf{x}$ , and therefore confidence procedure needs to be marginalized to  $\mathbf{x}_i$ . To find a  $(1 - \alpha)$  confidence interval for  $\mathbf{x}_i$  based on  $\mathbf{y}$ , just take the convex hull of all of the  $\mathbf{x}_i$  values in a  $(1 - \alpha)$  confidence set  $C(\mathbf{y})$ . We denote this confidence interval for  $\mathbf{x}_i$  as

$$C_i(\mathbf{y}) = \left\{ \min_{\mathbf{x}_i} C(\mathbf{y}), \max_{\mathbf{x}_i} C(\mathbf{y}) \right\}. \quad (4)$$

$C_i$  is typically conservative even if the original confidence procedure is exact. As we prefer exact confidence procedures to conservative ones, it is interesting that there is a special case which gives exact confidence intervals for  $\mathbf{x}_i$ .

Assume that  $\mathbf{F}(\cdot)$  is an affine function of  $\mathbf{x}$ , which we write as

$$\mathbf{F}(\mathbf{x}) = \mathbf{a} + \mathbf{K}\mathbf{x}, \quad (5)$$

where  $\mathbf{a}$  is the axis intercept of size  $m$  and  $\mathbf{K}$  is an  $m \times n$  Jacobian matrix. Treating  $\mathbf{a}$  and  $\mathbf{K}$  as given, this has the

form of a standard regression problem, which has a maximum likelihood estimate

$$\hat{\mathbf{x}} = \mathbf{S}\mathbf{K}^T\mathbf{S}_e^{-1}(\mathbf{y} - \mathbf{a}), \quad (6a)$$

where

$$\mathbf{S} = (\mathbf{K}^T\mathbf{S}_e^{-1}\mathbf{K})^{-1}. \quad (6b)$$

The sampling distribution of  $\hat{\mathbf{x}}$  is

$$\hat{\mathbf{x}}(\mathbf{y}) \sim \mathcal{N}(\mathbf{x}, \mathbf{S}).$$

This implies that there is an exact confidence procedure for the affine case,

$$C'(\mathbf{y}) = \left\{ \mathbf{x} : [\mathbf{x} - \hat{\mathbf{x}}]^T \mathbf{S}^{-1} [\mathbf{x} - \hat{\mathbf{x}}] \leq \chi_n^{-2}(1 - \alpha) \right\}. \quad (7)$$

We can prove that Equation (7) is the same confidence procedure as Equation (3) under Equation (5), by completing the square. This confidence procedure produces elliptical confidence sets.

Additionally, Equation (5) implies that there is exact confidence interval for each  $\mathbf{x}_i$ . The marginalization property of the Normal distribution implies that

$$\hat{\mathbf{x}}_i(\mathbf{y}) \sim \mathcal{N}(\mathbf{x}_i, \mathbf{S}_{ii}),$$

from which

$$C_i(\mathbf{y}) = \left\{ \mathbf{x}_i : \mathbf{x}_i \in \hat{\mathbf{x}}_i \pm z_{\alpha/2} \sqrt{\mathbf{S}_{ii}} \right\} \quad (8)$$

is an exact confidence interval, where  $z_{\alpha/2}$  is the quantile function of the Normal distribution.

To summarize this section, we have a variety of approaches. Under the statistical model alone, we have an exact confidence procedure for  $\mathbf{x}$ , given in Equation (3), but only a conservative confidence procedure for  $\mathbf{x}_i$ . But if  $\mathbf{F}(\cdot)$  has the affine form in Equation (5), then we also have an exact confidence procedure for  $\mathbf{x}_i$ , Equation (8). If  $\mathbf{F}(\cdot)$  is not affine, we still have the option of computing the Jacobian matrix at  $\hat{\mathbf{x}}$ , and using the affine form as a tangential approximation to  $\mathbf{F}(\cdot)$ . In this case, Equation (8) is an “approximately exact” confidence procedure for  $\mathbf{x}_i$ : its coverage would be exactly  $(1 - \alpha)$  were the tangential approximation correct, but its coverage deviates from  $(1 - \alpha)$  as the tangential approximation deviates from  $\mathbf{F}(\cdot)$ . The inappropriate use of the affine

approximation to  $\mathbf{F}(\cdot)$  creates a coverage error, in which the nominal coverage of at least  $(1 - \alpha)$  everywhere is not guaranteed, and may well be lower for some values of  $\mathbf{x}$ .

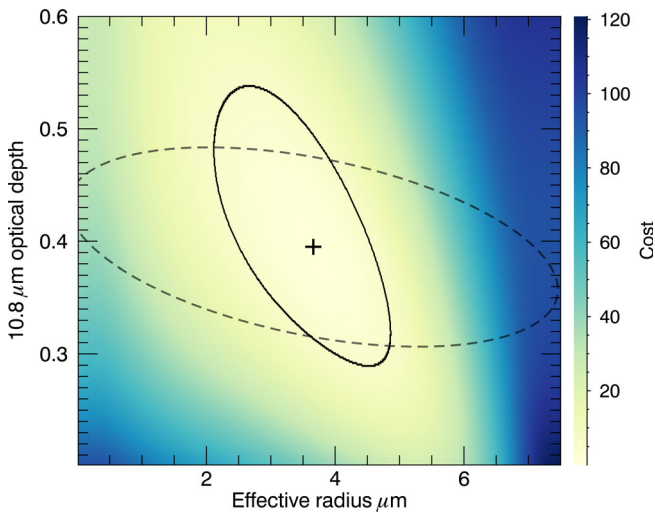
### 3 | COMPARISON OF UNCERTAINTY REGIONS

To compare the confidence procedures and intervals we use simple examples of retrieving two parameters in  $\mathbf{x}$ . The SEVIRI sensor on board the Meteosat Second Generation satellite provides the observations for both examples. SEVIRI has a 3 km nadir pixel resolution at the wavelengths of interest, which increase in size with zenith angle, and has a baseline repeat cycle of fifteen minutes.

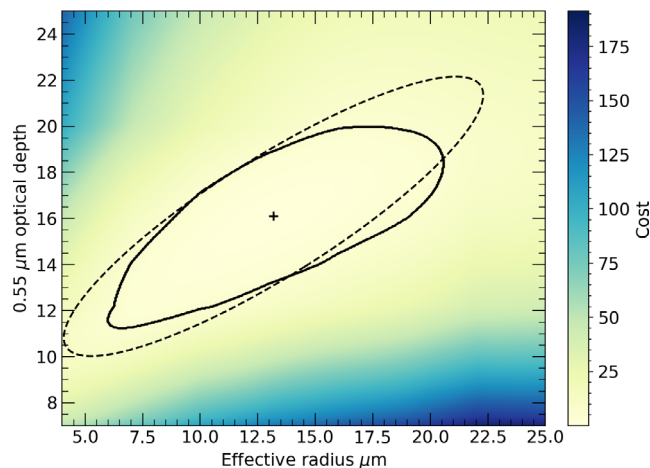
The first example retrieves the effective radius of a log-normal distribution of particles (with a fixed geometric standard deviation of 2.0) and optical depth at 10.8  $\mu\text{m}$ , from an observation of a volcanic ash cloud (with an assumed fixed altitude of 5.5 km) at 1800 UTC on 6 May during the 2010 eruption of Eyjafjallajökull located at 58.09°N 10.92°W. The observations are two channels centred at 10.8  $\mu\text{m}$  and 12.0  $\mu\text{m}$ , and a simple forward model for volcanic ash radiative transfer (see Francis *et al.*, 2012) using refractive indices for andesite (Pollack *et al.*, 1973). The radiative transfer program RTTOV (Hocking *et al.*, 2013) simulates the radiances using meteorology from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim Global Reanalysis data set (Dee *et al.*, 2011) on 60 model levels. The radiances are simulated with effective radius values between 0.1 and 10.0  $\mu\text{m}$  and 10.8  $\mu\text{m}$  optical depths between 0.1 and 1.0. Francis *et al.* (2012) provide the values for  $\mathbf{S}_e$ .

The second example retrieves the effective radius and optical depth, valid at a reference wavelength of 0.55  $\mu\text{m}$ , of a water cloud at 1300 UTC 31 October 2019 located at 12.95°S 1.86°E. The observations are the measured reflectance at 0.8 and 1.6  $\mu\text{m}$  and the Discrete Ordinates Radiative Transfer Program for a Multi-Layered Plane-Parallel Medium (DISORT, Stamnes *et al.*, 1988) simulates the reflectance values given the cloud properties and the relevant angles for the geometry of the problem. The reflectance values are simulated with effective radius values between 4 and 28  $\mu\text{m}$  and at a reference 0.55  $\mu\text{m}$  optical depth between 1 and 665. The combined standard deviation for the error in the measured and simulated observations is assumed to be 5% of the measured value.

Figure 1 shows the confidence region at level  $(1 - \alpha) = 0.95$  using Equation (3), and under the



**FIGURE 1** The “cost” for the effective radius and 10.8  $\mu\text{m}$  optical depth of a volcanic ash cloud from radiances measured by the SEVIRI sensor. The figure shows the 95% exact confidence region using Equation (3) (solid) and the confidence region under the assumption of an affine function using Equation (7) (dashed). The black cross indicates the maximum likelihood estimate



**FIGURE 2** The “cost” for the effective radius and 0.55  $\mu\text{m}$  optical depth of a water cloud from radiances measured by the SEVIRI sensor. The figure shows the 95% exact confidence region using Equation (3) (solid) and the confidence region under the assumption of an affine function using Equation (7) (dashed). The black cross indicates the maximum likelihood estimate

assumption of an affine function using Equation (7) for the volcanic ash cloud and Figure 2 shows the same for the water cloud. Figures 1 and 2 show that a locally linear, or affine, assumption using Equation (7) has a level set error when describing the uncertainty in nonlinear optimal estimation. Instead it is preferable to

use the exact conservative procedure Equation (3). For the given examples, the level set error using Equation (7) is greater for the volcanic ash cloud than the water cloud, although this will vary for each specific case. The level set error using Equation (7) will not be universal between retrievals of cloud type, nor between distinct sounding of the same cloud type. The level set error using Equation (7) instead relies on the suitability of assuming the simulated model is affine over the confidence interval. Transformation of variables, such as inferring the logarithm of variable, may reduce the coverage error; however the confidence region using Equation (7) is not transformation-invariant like the exact conservative procedure (3). This confidence procedure requires knowledge of the likelihood (or “cost”) to the point that Equation (3) can be evaluated. This may be computationally expensive for iterative procedures with a large parameter space. However, the exact confidence procedure (3) may prove especially useful where the entire parameter space has been forward modelled prior to or during the retrieval, such as the use of a look-up table (e.g., Nakajima and King, 1990; Wen and Rose, 1994) or problems with a small parameter space (e.g., Eyre and Menzel, 1989; Taylor *et al.*, 2019). In this case the exact confidence procedure negates the need to compute the Hessian matrix in Equation (7) under the assumption in Equation (5), decreasing the cost of computation, with the added benefit of a lower coverage error and transformation-invariant confidence regions.

## 4 | CONCLUSIONS

Those who quantify nonlinear optimal estimation uncertainty in cloud and aerosol remote sensing may wish to use an alternate approach to linear approximations if there is a desire to better quantify the region of uncertainty. For procedures where the computational cost of forward modelling the parameter space has already been done, and the maximum likelihood is found by calculating the likelihood for the whole parameter space, then it is both computationally favourable and has a lower coverage error to use the confidence procedure (3). The confidence set (3) for  $\mathbf{x}$  is exact, and does not rely on a linearization approximation such as  $\mathbf{y} \approx \mathbf{a} + \mathbf{K}\mathbf{x}$ , or possibly  $\mathbf{y} \approx \mathbf{a} + \mathbf{K}\mathbf{g}(\mathbf{x})$ , where  $\mathbf{g}(\cdot)$  is a transform used to linearize the problem. The coverage of a confidence set for  $\mathbf{x}$  derived using a linear approximation will not equal the nominal coverage. A simple way to find the error in the approximation, or to find an appropriate transform  $\mathbf{g}(\mathbf{x})$ ,



would be to compare it with an exact confidence set, such as that in Equation (3). If accuracy of coverage is required, the linearization approximations are redundant.

This procedure requires forward-simulating observations for sets of input variables, either on a regular or variable grid. The input variables which most closely map the simulated observations to the measured observations are the maximum likelihood estimates for each parameter. The confidence interval is evaluated at a level according to the confidence required using the chi-squared distribution with the degrees of freedom equal to the number of observations. For example, we may wish to produce a confidence region at the 95% level, where the retrieval consists of two observations. In this case, the input variables that evaluate to less than 5.991 (a constant following the chi-squared distribution with two degrees of freedom) using Equation (3) fall inside the confidence region. This is equivalent to taking all input variables where the “cost” of the retrieval evaluates to less than 5.991 in a less statistical approach. The minimum and maximum simulated value for each parameter that falls within this confidence region forms the confidence interval for this parameter at the required confidence level.

## ACKNOWLEDGEMENTS

Luke Western was supported by grant ref. NE/K007912/1 a NERC Met Office CASE studentship. We acknowledge EUMETSAT for the provision of the SEVIRI data.

## ORCID

Luke M. Western  <https://orcid.org/0000-0002-0043-711X>

## REFERENCES

- Benas, N., Meirink, J.F., Stengel, M. and Stammes, P. (2019) Sensitivity of liquid cloud optical thickness and effective radius retrievals to cloud bow and glory conditions using two SEVIRI imagers. *Atmospheric Measurement Techniques*, 12(5), 2863–2879, <https://doi.org/10.5194/amt-12-2863-2019>.
- Casella, G. and Berger, R.L. (2002) *Statistical Inference*, Vol. 2. Pacific Grove, CA: Duxbury.
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137 (656), 553–597, <https://doi.org/10.1002/qj.828>.
- Eyre, J.R. and Menzel, W.P. (1989) Retrieval of cloud parameters from satellite sounder data: a simulation study. *Journal of Applied Meteorology*, 28(4), 267–275, [https://doi.org/10.1175/1520-0450\(1989\)028<0267:ROCPFS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1989)028<0267:ROCPFS>2.0.CO;2).
- Francis, P.N., Cooke, M.C. and Saunders, R.W. (2012) Retrieval of physical properties of volcanic ash using Meteosat: a case study from the 2010 Eyjafjallajökull eruption. *Journal of Geophysical Research: Atmospheres*, 117(D20), 1984–2012, <https://doi.org/10.1029/2011JD016788>.
- Hocking, J., Rayer, P. J., Rundle, D., Saunders, R. W., Matricardi, M., Geer, A., Brunel, P. and Vidot, J. (2013). *RTTOV v11 Users Guide*. NWP-SAF Report. UK, Met Office.
- Mackie, S., Embury, O., Old, C., Merchant, C.J. and Francis, P. (2010) Generalized bayesian cloud detection for satellite imagery. part 1: Technique and validation for night-time imagery over land and sea. *International Journal of Remote Sensing*, 31(10), 2573–2594, <https://doi.org/10.1080/01431160903051703>.
- Mardia, K. V., Kent, J. T. and Bibby, J. M. (1980). *Multivariate Analysis*. London: Academic Press
- Morey, R.D., Hoekstra, R., Rouder, J.N., Lee, M.D. and Wagenmakers, E.-J. (2016) The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103–123, <https://doi.org/10.3758/s13423-015-0947-8>.
- Nakajima, T. and King, M.D. (1990) Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. part I: Theory. *Journal of the Atmospheric Sciences*, 47(15), 1878–1893, [https://doi.org/10.1175/1520-0469\(1990\)047<1878:DOTOTA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2).
- Pavolonis, M.J., Heidinger, A.K. and Sieglaff, J. (2013) Automated retrievals of volcanic ash and dust cloud properties from upwelling infrared measurements. *Journal of Geophysical Research: Atmospheres*, 118(3), 1436–1458, <https://doi.org/10.1002/jgrd.50173>.
- Peers, F., Francis, P., Fox, C., Abel, S.J., Szpek, K., Cotterell, M.I., Davies, N.W., Langridge, J.M., Meyer, K.G., Platnick, S.E. and Haywood, J.M. (2019) Observation of absorbing aerosols above clouds over the south-east atlantic ocean from the geostationary satellite SEVIRI–Part 1: Method description and sensitivity. *Atmospheric Chemistry and Physics*, 19(14), 9595–9611, <https://doi.org/10.5194/acp-19-9595-2019>.
- Pollack, J.B., Toon, O.B. and Khare, B.N. (1973) Optical properties of some terrestrial rocks and glasses. *Icarus*, 19(3), 372–389, [https://doi.org/10.1016/0019-1035\(73\)90115-2](https://doi.org/10.1016/0019-1035(73)90115-2).
- Poulsen, C.A., Siddans, R., Thomas, G.E., Sayer, A.M., Grainger, R. G., Campmany, E., Dean, S.M., Arnold, C. and Watts, P.D. (2012) Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR. *Atmospheric Measurement Techniques*, 5(8), 1889–1910, <https://doi.org/10.5194/amt-5-1889-2012>.
- Rodgers, C.D. (2000) *Inverse Methods for Atmospheric Sounding: Theory and Practice*, Vol. 2. Singapore: World Scientific.
- Stamnes, K., Tsay, S.-C., Wiscombe, W. and Jayaweera, K. (1988) Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics*, 27(12), 2502–2509, <https://doi.org/10.1364/AO.27.002502>.
- Taylor, I.A., Carboni, E., Ventress, L.J., Mather, T. and Grainger, R. G. (2019) An adaptation of the CO<sub>2</sub> slicing technique for the infrared atmospheric sounding interferometer to obtain the height of tropospheric volcanic ash clouds. *Atmospheric Measurement Techniques*, 12, 3853–3883, <https://doi.org/10.5194/amt-12-3853-2019>.
- Wen, S. and Rose, W.I. (1994) Retrieval of sizes and total masses of particles in volcanic clouds using AVHRR bands 4 and 5.

*Journal of Geophysical Research: Atmospheres*, 99(D3), 5421–5431, <https://doi.org/10.1029/93JD03340>.

Western, L.M., Rougier, J. and Watson, I.M. (2018) Decision theory-based detection of atmospheric natural hazards from satellite imagery using the example of volcanic ash. *Quarterly Journal of the Royal Meteorological Society*, 144(711), 581–587, <https://doi.org/10.1002/qj.3230>.

**How to cite this article:** Western LM, Rougier JC, Watson IM, Francis PN. Evaluating nonlinear maximum likelihood optimal estimation uncertainty in cloud and aerosol remote sensing. *Atmos Sci Lett.* 2020;e980. <https://doi.org/10.1002/asl.980>